Multi-task Agents and Incentives: the Case of Teaching and Research for University Professors *

Marta De Philippis†

January 2017

Abstract

This paper evaluates the behavioural responses of multitask agents to the provision of incentives skewed towards one task only. In particular it studies the case of strong research incentives for university professors and it analyzes their effects on the way university faculty members allocate effort between teaching and quantity and quality of research, and on the way they select into different types of universities. I first obtain different individual level measures of teaching and research performance. Then, I estimate a difference in difference model, exploiting a natural experiment that took place at Bocconi University, which heavily strengthened incentives towards research in 2005. I find evidence that teaching and research efforts are substitutable in the professors' cost function: the impact of research incentives is positive on research activity and negative on teaching performance. The effects are driven by career concerns rather than by the monetary incentives and are stronger for low ability researchers. Moreover, under the new incentive regime lower ability researchers tend to leave the university. Since I estimate that teaching and research ability are positively correlated, this implies that also bad teachers tend to leave the university. These results are consistent with a model of incentives where agents allocate effort between two substitute tasks and ability is multidimensional.

Keywords: multitasking, incentives, teaching.

*I thank Steve Pischke for very precious guidance, supervision and encouragement. I thank Marco Agliati, Oriana Bandiera, Marco Bertoni, Tito Boeri, Alessandra Casarico, Stephan Maurer, Michele Pellizzari, Giovanni Pica, Jesse Rothstein and Giulia Zane and participants to the LSE work in progress seminar, the LSE labour workshop, the second fRDB workshop and the XIII Brucchi Luchino workshop, the Padova University labour seminar for providing me with very useful comments and information. Finally I am indebted to Mariele Chiruli, Enrica Greggio, Erika Palazzo, Cherubino Profeta and Gianluca Tarasconi for precious help and information on the data.

†Bank of Italy and CEP, London School of Economics. e-mail address: M.de-philippis@lse.ac.uk

1 Introduction

The study of principal-agents relationships and the design of the optimal incentive provision systems have a long tradition in economics. A particularly complex and very common situation arises when agents have to allocate their time and effort among different tasks. In this case, the provision of incentives on one task only may distort multi-task agents' behaviour: individuals may respond by increasing effort in the activities subject to incentives, crowding out time and energy from other uses. This is especially the case if performance in other tasks is not easy to measure, and if there are no other reasons, such as social pressure or intrinsic motivation, to perform them in any case [Holmstrom and Milgrom, 1991, 1994, Brüggen and Moers, 2007, Fehr and Fischbacher, Bandiera et al., 2010, Benabou and Tirole, 2003, Prendergast, 2008].

While the theory related to multi-task agents is very well-developed, starting from the seminal work by Holmstrom and Milgrom [1991], empirical tests to the size and the sign of the behavioural responses predicted by this type of models are difficult to implement because of very heavy data requirements, first of all the need of an individual measure of performance for each task, which is often not easily observable. The empirical literature is therefore scarce and the actual economic cost of standard incentives for multitasks agents is still largely unknown. In practice, it depends on how the different tasks interact in the agent's production and cost functions and it is therefore specific to the actual tasks taken into consideration.

This paper analyzes one of the leading examples of multi-task agents: the case of university professors. Faculty members allocate time among many activities, mostly teaching and research. Incentives in most countries are however strongly skewed towards research: the 'Publish or Perish' paradigm is the most popular criterion for faculty hiring and promotion decisions in universities. This paper analyses the overall consequences of strong research incentives on teaching and research outcomes. It evaluates, first, the direct impact of research incentives on research performance itself; second, it studies the indirect effect of research incentives on teaching quality. Moreover, in order to understand the overall impact on teaching and research performance at the university level, it analyses how the composition of professors in terms of teaching and research skills changes under an incentive scheme more skewed towards research. Finally, by analysing the correlation between teaching and research skills for each professor, this paper discusses what may be the costs and benefits of separating teaching and research careers for university faculty.

Using a standard model of incentives where agents allocate effort between two different tasks and ability is multidimensional, I show that the way stronger research incentives affect research activity, teaching performance and the selection of professors in the university depends on two main parameters: on whether teaching and research are substitute or complement in the professors' cost function and on whether teaching and research skills are positively or negatively correlated (i.e. whether good researchers are also good teachers). I then estimate the sign of these parameters. I overcome many of the standard identification issues by studying the case of Bocconi University, an Italian private institution of tertiary education based in Milan. Its institutional setting provides a unique opportunity to test the effect of research-oriented incentives on teachers' performance in multiple activities and the overall effect on the university teaching and research outcomes, for three reasons. First, I can construct a measure of teaching performance using a value added approach, that is the standard one used to evaluate teachers in primary or secondary schools [Rothstein, 2010, Rockoff, 2004, Aaronson et al., 2007, Rivkin et al., 2005]. It is usually difficult to apply this method for university professors because students self-select into courses, exams and teachers and the usual assumption that - conditional on previous test scores - allocation of students to teachers is random is not credible in the university context.[1] Bocconi students are instead randomly assigned to teachers in each academic year: within a degree program, if the number of enrolled students requires it, students are randomly split in different classes, each of which taught by a different lecturer, while the exam, the syllabus and the classrooms are identical for all students. Therefore, once I include an exam script fixed effect, I can use the average class grade of students taught by different professors teaching the same course as a proxy for teaching quality.[2] Second, Bocconi sharply changed its faculty's incentive regime in 2005, shifting the focus explicitly towards research, by strengthening research requirements for promotion decisions and by introducing monetary incentives based on

[1] The only papers that can apply the value added method in the university context analyze the case of Bocconi university [Braga et al., 2014, 2016] or the case of the U.S. Air Force Academy Carrell and West [2010] where the institutional setting implies student-teacher randomization.

[2] In particular I estimate time varying teachers fixed effects, controlling for yearly shocks at the course level-such as shocks to the exam papers or to the syllabus. In principle, once I include the course fixed effects, I do not need to control for students' characteristics such as previous test scores, because of randomization of students across classes within the same course.

quality and quantity of publications. Third, the large heterogeneity of Bocconi teaching contracts provides a natural control group: many teachers act only as external teaching faculty. They have the same teaching responsibilities but are not subject to Bocconi changes in promotion strategy and research incentives.

This paper therefore estimates a difference in difference equation, evaluating teachers' performance before and after 2005 and using external teachers as control group. The robustness of the results is confirmed by using two alternative control groups: faculty members who became tenured just before 2005, and are therefore not exposed any more to career concerns but only to the change in monetary incentives, and faculty members of another Italian university (Bologna), very similar to Bocconi in terms of quality and quantity of research, but not subject to any change in incentives.[3]

My main results are as follows. First, the new incentive regime improved both the quality and the quantity of published papers. After the change in the incentive scheme Bocconi faculty members started to publish, on average, more papers than before, by 9% of a standard deviation. Moreover the effect is mostly driven by young faculty members, whose career concerns are stronger since they are not tenured yet. This result is in line with the literature on piece rate incentives [Lazear, 2000, Bandiera et al., 2007]. Second, the introduction of incentives towards research had a negative impact on teaching performance, as measured by time-varying teacher fixed effect. In particular teaching quality decreased by 7% of a standard deviation under the new incentive regime. The effect is, again, mostly driven by young faculty members and more negative for students at the bottom of the ability distribution. Combining the two estimates on teaching and research I find that, overall, one extra publication reduces teaching quality by one third of a standard deviation. This suggests that teaching and research are substitutes, not complement in the teachers' cost functions, at least for the type of courses I am considering. Third, there is evidence of some positive selection effects: the new incentive scheme induced low ability researcher to leave the university, thus increasing the average quality of research at the university. Forth, I document that teaching and research skills are positively correlated: if a university manages to attract/maintain good researchers, it will

[3] This second strategy can only be applied to Research outcomes, because there is no information on teaching performance for the university of Bologna. I chose Bologna, because in terms on quality of research as evaluated by the Italian Institute of University Research Evaluation (ANVUR) it is the most similar to Bocconi University, in terms of dimension of the department and quality of the research outcome between 2004-2010. www.anvur.org/rapporto/files/Area13/VQR2004-2010_Area13_Tabelle.pdf

also attract/maintain good teachers. The overall effect on teaching quality is therefore ambiguous: on the one side, since teaching and research effort are substitutes, teaching quality of incumbents decreases, on the other side the policy pushes away the worst researchers and, since research and teaching skills are positively correlated, also the worst teachers.

This paper fits into the literature that investigates behavioural responses to incentives, in particular in the context of multi-task agents. As mentioned before, there is little empirical evidence of the impact of incentives for multi-task agents and on the actual cost of providing incentives on one task only, mostly because of limitations in the data and of the difficulty in simultaneously measuring performance in many tasks, for instance because it is not observable or it is difficult to disentangle the individual contribution to the final outcome. Few exceptions are Dumont et al. [2008], Feng Lu [2012], Hong et al. [2013], Johnson and Reiley [forthcoming]. In the education literature Jacob [2005], Fryer and Holden [2012] analyze the impact of accountability policies on test-specific skills and students' effort in high-stake versus low-stake exams.

My paper contributes to the incentive literature, first, by providing a well-identified estimate of how multiple tasks interact in the agents' cost function. While most of the existing papers look at the quality-quantity trade off of performing the same activity, I analyse the effect on the performance in two different activities, when it is not clear a priori whether the tasks are substitute or complement in the agents' cost function. Second, to my knowledge this is the first paper that combines estimates of the effort substitution effect with an analysis of how multi-task agents sort in different types of firms, depending on their incentive schemes. This is key in order to evaluate the overall effect for the principal of different incentive schemes. Sorting effects may be very relevant and they may countervail the direct effort substitution effect so to revert the sign of the impact of changes in the incentive scheme. Third, I am able to disentangle the pure effect of monetary incentives from the effect generated also by career concerns: understanding the main drivers behind different responses is extremely useful in order to efficiently design incentives in other contexts or settings.

My paper is also related to the education literature on teachers' contracts and incentives. Some papers evaluate the effect of teaching contracts on teaching performances [Figlio et al., 2015, Bettinger and Terry, 2010, Figlio and Kenny, 2007] and find that students learn more from non-tenure line professors. Since non-tenure line faculty is less focussed on research, this may suggest that these results are driven by differences in teachers' incentive schemes. Still, it is impossible from these analyses to disentangle whether the effect they find is actually driven, instead, by selection into non-tenure line jobs. Some other papers look more directly at the trade-off between teaching and research, by analyzing the effect of increased teaching incentives on research and teaching outcomes. Brickley and Zimmerman [2001] use a single difference strategy to study the consequences of the introduction of teaching performance incentives at the University of Rochester Business School. The authors find a substantial and almost immediate jump in teaching ratings, measured by students' evaluations, and a corresponding decline in research output. Payne and Roberts [2010] analyze this same issue but using between, not within, university variation. They exploit US state variation in the adoption of teaching performance measures and find that research activity decreased in quantity but improved in quality in non-flagship universities.

This paper contributes to the education literature in two ways. First, it is the only one, to my knowledge, to test the other side of the relationship between teaching and research: the effect of strong research incentives. This type of analysis is crucial given the extremely wide adoption of research incentives in universities. Moreover, it is likely that the extent of effort reallocation generated by research incentives is larger than for teaching incentives because teaching effort is more difficult to measure and monitor and peer pressure on excellence in teaching is much weaker than in research. Moreover, this paper provides the first piece of evidence about the sign of the correlation between teaching and research skills. The positive correlation between teaching and research skills has important implications for the design of professors' incentives and hiring schemes. For example, policies aimed at increasing teachers' specialization that propose to dedicate part of the faculty exclusively to teaching and part of it exclusively to research, should take into consideration that there is substantial overlap between good researchers and good teachers.

The structure of the paper is as follows. Section 2 provides a simple conceptual framework that rationalizes expected results; Section 3 describes the data; Section 4 outlines the identification strategy; Section 5 presents my empirical results and Section 6 shows how my results are robust to alternative control groups. Finally, Section 7 briefly characterizes the policy implications of my results and concludes.

2 Conceptual framework

This section presents a very simple framework aimed at organizing and rationalizing the expected findings. The working of the model in the spirit of Holmstrom and Milgrom [1991] and it is similar to the model presented by Fryer and Holden [2012].[4]

An agent, upon accepting the contract, takes two non-verifiable actions e_r and e_t, which I call research and teaching effort respectively. Each action takes values in $\mathbb{R}_+$ and generates a performance measure $m_i = \alpha_i e_i$ where $i = r, t$ and α_i is unknown to the principal. I refer at α_i as the type of the agent on task i (her ability level).

I assume that the principal offers a linear incentive scheme of the form $x = s + b_r m_r + b_t m_t$, where b_r and b_t represent the variable component of the wage, related to the research and teaching performance, and s is the fixed part of the wage.[5] If the agent accepts, she makes her effort choices, the performance measure is realized and the principal pays the agent accordingly.

I also assume that the agent's preferences can be represented by the following CARA utility function:

$$u(x, e) = -exp[-\eta(x - \frac{1}{2}(e_r^2 + e_t^2) - \delta e_r e_t)] \tag{1}$$

where x is the monetary payment and δ is the degree of substitutability between the tasks r and t in the cost function. Let $\underline{U}$ be the agent's outside option if he does not work. Moreover, I assume that there is a minimum teaching performance $\underline{m}_t$ and research performance $\underline{m}_r$ required by the university.

The agent therefore maximizes utility with respect to e_r and e_t, subject to the participation constraints ($u(x, e) > \underline{U}$ and $m_r > \underline{m}_r$ and $m_t > \underline{m}_t$). Note that when $m_r^* < \underline{m}_r$ or $m_t^* < \underline{m}_t$ each individual will choose whether to stay and exert effort level $\underline{e}$ (such that $\underline{m}_r = \underline{e}_r \alpha_r$) or to leave and stop working, depending on whether $U(x_{\underline{m}}, e_{\underline{m}})$ is larger or smaller than $\underline{U}$.[6] If it is smaller,

[4] I will not model why the university decided to increase research incentives, i.e. I do not make assumptions on the university objective function, I only analyze what are the agents' responses to an increase in research incentive, in the spirit of Lazear [2000].

[5] This payment scheme, which entails additive separability between teaching and research, reflects the way the two activities are rewarded by Bocconi University: in the description on how research and teaching incentives are provided there is no mention and no relationship with the way professors performed in the other task.

[6] $e_{\underline{m}}$ and $U(x_{\underline{m}}, e_{\underline{m}})$ are respectively the effort an agent needs to exert in order to obtain $\underline{m}$ and the utility level when $\underline{m}_r$ and or $\underline{m}_t$ are binding.

she will decide to leave (or be fired). Otherwise, she will be induced to exert more effort, even if very costly, in order to stay in the university.

The new Bocconi's incentive scheme, which I will describe in more detail in Section 4, implied an increase in b_r, the monetary return to research activity, and in $\underline{m}_r$, the minimum research performance required, but only for professors not tenured yet. Changes in b_r act mostly on the intensive margin (the amount of research effort to exert), changes in $\underline{m}_r$ instead mostly affect decisions on the extensive margin (whether to stay in the university or to leave).

2.1 Effects on teaching and research performances

This section shows what happens to m_r^* and m_t^* (and therefore e_r^* and e_t^*) if the university increases b_r and $\underline{m}_r$ (to $\underline{m}'_r$) and professors remain in the university.

In the Appendix I solve the model (for internal solutions) and I show that the equilibrium effort level is:

$$e_r^* = \frac{b_r\alpha_r - \delta b_t\alpha_t}{1-\delta^2}; \quad e_t^* = \frac{b_t\alpha_t - \delta b_r\alpha_r}{1-\delta^2} \tag{2}$$

It is clear that e_r^* increases if b_r increases, while the sign of the derivative of e_t^* with respect to b_r depends on the sign of δ.

Proposition 1 *An increase of b_r, the marginal return on research performance, leads to an increase in e_r.*

The response of e_t depends on the value of δ: $\begin{cases} \frac{\partial e_t}{\partial b_r} < 0 & if\,\delta > 0 \text{ } (e_r \text{ and } e_t \text{ substitute}) \\ \frac{\partial e_t}{\partial b_r} > 0 & if\,\delta < 0 \text{ } (e_r \text{ and } e_t \text{ complement}) \end{cases}$

The policy, moreover, increased $\underline{m}_r$.

Proposition 2 *When $m_r^* > \underline{m'}_r$: an increase in $\underline{m}_r$ does not have any effect.*

When $m_r^* < \underline{m'}_r$ and $U(x_{\underline{m'}_r}, e_{\underline{m'}_r}) > \underline{U}$, $\underline{m}_r$ is binding and professors exert $e_{\underline{m}_r,r}$ even if above their optimal level.

Where $m_r^* = \alpha_r e_r^*$; $U(x_{\underline{m'}_r}, e_{\underline{m'}_r})$ is the utility level achieved when research outcome $m_r = \underline{m'}_r$ and $\underline{U}$ is utility from leisure.

2.2 Effects on the composition of faculty members

Whether agents will decide to continue working under the new regime or to leave, depends on $\underline{U}$, the utility provided by leisure, on $U(x_{\underline{m'}_r}, e_{\underline{m'}_r})$, the utility provided by achieving the new minimum level of research and teaching performance in order to stay at the university and on $U(x, e*)$, the utility provided by optimizing research and teaching efforts, without constraints.

Increases in b_r, do not have any effect on the decision to continue working or to leave the university because, at most, the agents will not change their behaviour. Increases in $\underline{m}_r$, instead, may have effects on the decision to leave the university.

Proposition 3 *If $\underline{U} > max \quad \{U(x_{\underline{m'}_r}, e_{\underline{m'}_r}); U(x, e*)|_{m*>\underline{m'}_r}\}$, professors will leave the university and enjoy utility $\underline{U}$*

Therefore, overall, for individuals whose $m_r^* > \underline{m}'_r$, the effect of the policy comes entirely from variations in b_r and therefore from evaluating the sign of the derivatives of e_r^* and e_t^* with respect to b_r.

For individuals whose $m_r^* < \underline{m}'_r$, the effect depends on whether $U(x_{\underline{m}}, e_{\underline{m}})$ under the new $\underline{m}'_r$ is larger or smaller than $\underline{U}$. If it is smaller they will decide to leave and exert no effort. Otherwise, they will be induced to exert more research effort (in order to reach $\underline{m}'_r$), even if very costly, and stay in the university.

I now evaluate how this effect varies by agent's ability. Since $\frac{\partial e_r}{\partial \alpha_r}|_{m_r=\overline{m}} < 0$, i.e. research effort is more costly for low α_r individuals, an increase in research incentives will be much more beneficial for high ability researchers. For these researchers, indeed, the marginal benefit of an extra unit of effort is much higher. Instead, those more likely to leave because of an increase in $\underline{m}_r$ are low ability researchers.

Proposition 4 *When $m_r^* > \underline{m}_r$: an increase in b_r, leads to a larger increase in e_r and reduction e_t for individuals with high α_r. For low α_r agents, it is more likely that $\underline{m}'_r$ is binding, and that they are induced to leave.*

The predicted response of stronger b_r along the distribution of α_r is therefore that: (i) for research the effect is non-linear. Very low ability researchers will leave the university; of those staying, the lowest ability ones (those whose $m_r^* < \underline{m}'_r$) will increase effort on research in order to reach $\underline{m}'_r$; the others (those whose $m_r^* > \underline{m}'_r$), will increase e_r proportionally with their ability α_r, because for high α_r individuals the marginal return of an extra unit of effort is higher; (ii) for teaching, the effort substitution effect simulates the response on the research side: it is stronger for agents whose response on the research side is stronger (as long as $\delta > 0$).

3 Data and descriptive statistics

3.1 Students

This paper uses the administrative records of individual students and teachers from Bocconi University, an Italian private institution of tertiary education based in Milan. Bocconi offers degree programs in Economics, Management and Law. I only consider compulsory undergraduate courses between 2001 and 2011. My sample includes around 700 teachers and 30,000 students, who take on average 20 compulsory exams over the 3 years of study.

My data cover in detail the entire academic history of students, including their basic demographics (gender, place of residence and place of birth), high school leaving grades as well as high school type (whether the high school focusses on humanities, on sciences or technical/vocational subjects). Information is also provided on the grades in each single exam together with the date when the exams were sat. Moreover, I have access to the random class identifiers of students, which allow me to match students with teachers.[7]

Table 1 reports descriptive statistics for students. Most students are Italian, one fourth is from Milan. They are positively selected among the population of high school graduates: the average high school final grade is very high (0.9 out of a maximum of 1[8]). On average there are 5 classes per course, of about 110 students each, and 20 compulsory undergraduate courses per year. Each student sits on average 7 exams per year. The degree program in Management is the one with the

[7]About 3% of students did not sit the exam in the academic year they were supposed to and are randomly allocated to a new class. Since the records on the initial class allocation are overwritten in the administrative database, I include them in the new class, controlling for a dummy that signals if the student took the exam in a different year from what expected.

[8]Given that I know the maximum final high school grade each foreign student can take, I standardize high school final grades of foreign students to be between 0.6 and 1, so that they are comparable with grades of Italian students.

highest number of classes (7 on average).

3.2 Teachers

Together with data on students, I have access to administrative data on Bocconi faculty. In particular, I have information on teachers' demographics (date of birth, gender, full name), type of contract, department of affiliation and number of teaching hours in each course and class. I am therefore able to match students with teachers.

I classify each teacher as internal or external. Table B.1 lists all different teaching contracts available at Bocconi over the years I consider and the way I group them into five categories: assistant professors-junior researchers, associate professors, full professors, non academics and professors from other universities. I define teachers in the first three categories as internal, treated by incentives, and teachers in the last two categories as external, my control group. I exclude lecturers, see Section 4.2

Column (1) of Table 2 reports descriptives for internal teachers, column (2) for external teachers and columns (3) shows the difference between the two groups. In total, in my sample, I observe 681 teachers for 5 years on average. Internal teachers tend to be slightly older and to teach more hours at Bocconi. Most teachers are hired by the Management or Economics department. Finally, based on the data from 2005, one year before the change in incentives, internal professors represent about 70% of the sample.

3.3 Students-teachers randomization

The randomization of students to classes (taught by different teachers) is performed every year via a simple random algorithm that assigns a class identifier to each student, within each course.[9][10] Table B.2 provides evidence that teachers were actually randomized to students. Following Braga

[9]The university administration adopted the policy of repeating the randomization for each course with the explicit purpose of encouraging wide interactions among the students.

[10]In practice, students enrolled in a certain degree program are randomly assigned each year to a different class. Students in each class attend the same set of courses, but are taught by different professors. Sometimes the same course is taken, however, by more than one degree course (in different classes). Therefore, since randomization takes place within a degree program, I include fixed effects for the full interaction of year, courses (that absorb shocks to the exam script) and degree programs (that allow me to exploit the randomization of students to teachers).

et al. [2014, 2016], I show results of a regression of class (student) average characteristics on teachers' characteristics, controlling for fixed effects on the full interaction of courses, academic years and degree programs.[11] The null hypothesis under consideration is the joint significance of the coefficients on teacher characteristics. The F statistics are always very low, suggesting there is no significant correlation between students' and teachers' characteristics.

3.4 Publications

I collect publication data from the Web of Science website. In particular, I count professors' yearly publications in the fields categorized by Web of Science as 'business', 'maths' and 'economics'. Since for less recent years the Web of Science database only reports the author's first name initial and not the full name, I run a search only using the authors' first name initial, together with their surname.

Moreover, through the use of a web scraping program which makes automatic searches (one for each year/professor combination) from the Google Scholar website, I collect information on working papers. I restrict my research on the Google Scholar website to the following fields: 'social sciences, arts, and humanities' and 'business, administration, finance, and economics'. Since in this case, data on full names are available for all years, I look for full names.[12]

4 Empirical strategy

This section develops my empirical strategy, aimed at estimating the causal effect of increasing incentives towards publishing on teaching and research performance.

I use administrative data from Bocconi university archives to estimate two Difference-in-Difference models, one for teaching and one for research, exploiting the sharp change in Bocconi research incentives and using external faculty as control group.

I begin this Section by describing in more details the reform in Bocconi's incentives regime announced in July 2005 (Subsection 4.1). Subsections 4.2 and 4.3 present my empirical model for

[11]This is the level at which randomization takes place

[12]This procedure does not eliminate the possibility that the same working paper is counted more than once, if published in two different versions. However, this is still a measure of the effort one puts in that specific research. Moreover, this measure also contains the published version of the working papers. Accessed in December 2011.

the evaluation of the effect on teaching and on research activity. Finally, Subsection 4.4 describes how I estimate selection effects.

4.1 The new incentive policy

In 2005, Bocconi University unexpectedly announced the adoption of a new policy of hirings and promotions. The Board of Directors called for the Rector to make Bocconi one of the top five universities in Europe. As a consequence, the old hiring and promotion strategies, mainly based on national competitions and seniority, were replaced with new practices based on international standards. Since then[13], an independent committee, composed of faculty members from all disciplines, has been in charge of recruiting and promotions. Decisions have been centralized at the university level, making exceptions impossible. The importance of research outcomes in promotion decisions was clearly stated in all internal faculty contracts.

The goals of the New Strategic Plan, as announced in July 2005, were the following: (i) improving the systems to evaluate research produced by each professor (through the creation of an independent evaluation committee and the internationalization of evaluation criteria); (ii) adopting clear incentives on research (both monetary[14] and career-based); (iii) creating mechanisms to "attract and keep the best researchers worldwide".

The focus switched explicitly towards research, tenure decisions started to be based almost entirely on scientific productivity and the requirements on quantity and quality of research started to be much tighter.

4.2 Research performance

I first evaluate whether incentives on publishing have an impact on research quality or quantity.

I use three different measures of research performance: (i) the number of publications as collected from Web of Science; (ii) a proxy of the index actually used by Bocconi to evaluate researchers

[13]The actual implementation of the policy was in 2007, but throughout the analysis I will consider the year of the announcement, 2005, as the treatment year. Be aware that the full effect will be in place starting from 2007.

[14]Even if previously anticipated, Bocconi started to actually provide monetary incentives to its internal faculty in the academic year 2008. In particular there are three types of incentives: (i) the possibility of getting "research profile", with less teaching duties; (ii) research premia that depend on the number and the quality of publications; (iii) research funds, given to everybody who has reached a minimum level of research productivity in the previous two years. Publications were weighted depending on the quality of the journal)

(which is computed as the sum of the number of articles published by each professor, weighted by the quality of the journals as classified by Bocconi[15], divided by the number of co-authors) and (iii) the number of working papers and published papers from Google Scholar.

I then implement a Difference-in-Difference model by estimating the following equation for the years between 2001 and 2010:

$$pub_{py} = \theta_y + \theta_p + \gamma_{res}(internal_p * post2006_y) + \gamma_4 X_{py} + \eta_{py} \tag{3}$$

where pub_{py} are publications of professor p in year y; $internal_p$ is the internal status (in 2005); θ_y are time fixed effects; θ_p are teacher fixed effects; X_{py} are teacher characteristics (age, age squared) and η_{py} is the error term. I cluster standard errors by professor.

For sake of consistency, I include only teachers who were teaching classes I can use to estimate the teaching equation (see below equation 5).[16] Moreover, in order to exclude endogenous status switches from internal to external or vice versa after the introduction of the policy, I classify teachers as internal if they were internal in 2005, before the change in promotion strategy. In my robustness checks (Section 5.4) I check my results are not driven by this choice, by running the same analysis using contemporaneous status instead of status before 2006 as treatment, therefore including endogenous 'switches' in the effect. Furthermore, I drop internal lecturers. Lecturers are internal professors (i.e. have a full time contract with Bocconi) but with only teaching duties.[17] On one side, monetary research incentives are not provided to lecturers but, on the other side, the way lectureship decisions are taken has probably changed after 2006. They therefore do not represent a good control group. In a robustness check (Section 5.4), I include lecturers and interact them with the treatment. Finally, in order to use the same sample of teachers as in the analysis on teaching, I do not consider law professors and law courses: law's exams are usually oral exams so the set of

[15] Bocconi divides journals into 3 categories: A+ journals (i.e. Econometrica), to which it assigns a weight of 15; A journals (i.e. Economic Journal), weighted 7; B journals (i.e. Economic Letters), weighted 3. I classified journals using the list valid for the year 2007, available upon request.

[16] The difference in the number of observations between Table 4 and 7 is given by those teachers who were teaching more than one class per year or by the fact that some teachers do not teach compulsory undergraduate courses all years, but I still include those year observations in my analysis, for consistency over time.

[17] The difference between the position of lecturers and assistant/full professors is clear from what is stated in their contracts. The contract for assistant professors states "responsibilities include teaching and, most importantly, productivity in research". The contract for lecturers, instead, states that only teaching duties are expected from lecturers. Research activity is not even mentioned.

questions is not the same for all students. It is therefore difficult to use average grade as a measure of teaching quality.

This strategy relies on several identifying assumptions. First, the assumptions that the trend in research performance for internal and external professors was parallel before 2006. Second, the assumption that there are not other contemporaneous changes, in 2006, that affected differently internal and external teachers. Finally, in order to interpret the results, it is important to explore the presence of spillover effects, and how and whether they differently affected internal and external teachers.

In Section 5 I will test for the presence of parallel trends.

For what concerns the presence of other contemporaneous policies, Section 5.4 discusses and discards the possibility that my results are affected by a different response of internal and external teachers to a policy, implemented by Bocconi also in 2006, that required teachers to standardize their grades along a common distribution. Moreover, it is unlikely that my results are affected by contemporaneous changes taking place in other universities. While one may worry that other institution may change their own incentive schemes either in response to the change at Bocconi or because they were all exposed to a common shock - like new recommendations from the ministry of education or changes in the national law - external professors come from a broad variety of alternative universities and institutions, of rather large dimensions, and Bocconi implemented this new policy on its own initiative, not following national policies.

Finally, Section 5 will investigate the presence of spillover effects, like changes in the probability of co-authoring with internal professors or with new hires. On the one side spillover effects are one potential mechanism that amplifies my result: if the new incentive scheme attracted very good researchers or induced professors to put more effort into research and this has positive spillover effects on the rest of the internal faculty, it amplifies the effect of research incentives. On the other side, these positive spillover effects may influence external teachers as well as internal ones. For instance, if internal professors change their publication effort, external professors coauthoring with them might also change their research output. This type of spillovers would lead me to underestimate my effects, and it is relevant to understand whether they exist and in which direction they go.

4.3 Teaching performance

Second, I estimate my empirical model for the effect on teaching. I do it in two steps.

I use the average students grade in a class, controlling for yearly shocks at the exam paper, as a proxy for teaching quality. Students taking the same course are all taught the same syllabus and are all examined on the same questions, independently of the class to which they are (randomly) assigned. Some variations in the material and in the exam across degree programs are allowed (this is why I correct for the full interaction of courses, degree programs and years). Usually a senior member of the faculty acts as the course coordinator: he establishes the material to teach, manages possible complications and prepares the exam paper. Grading is instead generally delegated to the individual teachers, who typically are supported in the marking by teaching assistants.

The first step of my teaching analysis uses micro data from the student academic curriculum database and it is aimed at computing the average grade at the class level, conditional on students' high school final score and demographics.[18] I estimate the following equation:

$$grade_{ipcy} = \beta_0 + \beta_1 HSgrade_i + \beta_2 X_i + \alpha_{pcy} + u_{ipcy} \tag{4}$$

where $grade_{ipcy}$ is the grade obtained by student i, with teacher p[19], in year y, in course c (standardized at the course-year level to have mean 0 and standard deviation 1); $HSgrade_i$ is student i high school final grade; X_i are the students' individual characteristics (gender, age, whether Italian, whether from Milan, type of high school attended); u_{ipcy} is the error term. α_{pcy}, the year specific teacher fixed effect, is my parameter of interest.

The second step evaluates how the time-varying teacher fixed effects α_{ptc} evolve over time, in response to the change in incentive regime. I implement the same Difference-in-Difference estimation as in Section 4.2, changing the dependent variable. In particular, I estimate the following equation:

$$\widehat{\alpha_{pcy}} = \delta_p + \delta_{cy} + \gamma_{teach}(internal_p * post2006_y) + \gamma_2 X_{py} + \epsilon_{pcy} \tag{5}$$

[18] To reduce computational burden, I exploit randomization of students to teachers and I do not include students fixed effects.

[19] Since in around 40% of the cases more than one professor teaches the same class the actual meaning of p in this first case is the "professor mix" of the class.

where $internal_p$ is a dummy equal to one if the professor was internal before the change in incentives; $post2006_y$ is a post reform dummy; δ_p are teacher fixed effects;[20] δ_{cy} are fixed effects for the full interaction between academic years, courses and degree programs[21]; X_{py} are time-varying professor characteristics (age, age squared, experience in teaching undergraduate courses in Bocconi) and ϵ_{pcy} is the error term. I cluster standard errors by professor.
γ_{teach} quantifies the change in teaching performance of incumbent professors under the new incentive scheme.

The economics literature usually measures teacher quality by estimating a teacher fixed effect in equation 4. Here I allow teacher effects to vary over time and I analyze how they change in response to the positive shock in research activity. I overcome many of the standard identification problems because: (i) I eliminate concerns related to time constant factors by including teachers' fixed effects in my regressions: I only analyse how teaching performance evolves over time; (ii) I eliminate concerns related to time varying endogenous matching of students to teachers thanks to the randomization of students to teachers. There is a debate (Rothstein [2010, 2009], Ishii and Rivkin [2009], Kane and Staiger [2008], Chetty et al. [2014]) about whether value-added models perform weakly in the absence of randomization, because in this case teachers fixed effects may also identify endogenous matching between teachers and students. Results are mixed.However, endogenous matching is likely to be much stronger in the university context, where students self-select into courses and therefore teachers.

Finally, I also estimate the same effect running the analysis directly at the student level, as follows:

$$grade_{ipcy} = \zeta_p + \zeta_{cy} + \zeta_{teach}(internal_p * post2006_y) + \zeta_2 X_{ipy} + v_{ipcy} \qquad (6)$$

where all the variables are defined as before and v_{ipcy} is the error term.
While my preferred specification is the estimation of equation 5, because it is more easily interpretable in terms of changes in teaching quality at the professor level, this last specification will

[20] Notice that in this case p represents a single teacher. Therefore if a class was taught by multiple teachers I impute the (unique) class fixed effect to both teachers.

[21] Courses may have the same code but programs and exams may be different for different degree programs. Interacting also with degree programs allows me to exploit variation across teachers' performance when syllabus and exam papers are exactly the same (and over which the randomization of students to teachers takes place).

allow me to evaluate how the main effect is heterogeneous with respect to students' characteristics, in particular with respect to students' previous test scores, measured by their final high school grade.

The identifying assumption below equations 5 and 6 are the same discussed in section 4.2 for the results on research.

4.4 Effects on the composition of faculty members

To have a complete picture of the overall effect of the change in incentives on research and teaching quality, I analyse how the composition of internal professors changed after the new regime was introduced. As shown in Section 2, the change in minimum research requirements should push low ability researchers away. Whether this translates into maintaining also better teachers, it will depend on how teaching and research skills are correlated.

I analyse selection effects in two ways: first, I compare estimates with and without professors' fixed effects; second, I obtain direct estimates of the underlying teaching and research abilities and I analyse how the ability composition of teachers varies over time, looking both at teachers sorting in and sorting out.

In order to analyse sorting patterns, I estimate teaching and research skills as follow:
For teaching:

$$\widehat{\alpha_{pcy}} = \theta_p^t + \delta_{cy} + \gamma_2 Q_{py} + \epsilon_{pcy} \tag{7}$$

where α_{pcy} is the conditional average grade of students taught by professor p, teaching course c in year y; δ_{cy} are fixed effects for every course-degree program-year combination; Q_{py} are professors' time varying characteristics (age, age squared, years of experience at Bocconi); ϵ_{pcy} is an error term. Finally, θ_p^t, the professor fixed effects, is my proxy for underlying teaching skills.

Analogously, for research:

$$pub_{py} = \theta_p^r + \zeta_y + \zeta_2 Q_{py} + \eta_{py} \tag{8}$$

where pub_{py} is the number of papers published by professor p in year y; ζ_y are year fixed effects, that absorb any possible time trend in how difficult it is to publish papers over time; Q_{py} are professor characteristics (age, age squared and their department of affiliation); η_{py} is an error term. Again,

θ_p^r, the professor fixed effects, are my estimate of underlying research skills.

One first concern is that, since incentives are muted under the new scheme, it is not clear whether fixed effects based on teaching or research productivity after 2006 are a good proxy for ability. This would imply one should only use fixed effects evaluated before 2006. However, it would be impossible to test whether the new policy managed to attract more skilled professors, since I would not be able to estimate a teacher fixed effect for faculty members who entered after 2006, under the new incentive regime. In Figure 1, I follow Lazear [2000] and I show that, for professors who were teaching also before 2006, there is a strong positive correlation between fixed effects evaluated in the period before 2006 and those estimated for the period after 2006. Whenever it is possible (for sorting out effects), I will run my regressions also using fixed effects estimated on the pre-2006 period only.

For the analysis on selection of professors, I do not run a proper difference-in-difference strategy because, since I do not know the entire employment history of external teachers, I cannot use them as control group.[22] For internal professors, instead, I know exactly their year of entry, every change in their contracts and their year of exit, including the reason for leaving. Moreover, it is very unlikely that external teachers represent a good control group for the analysis on sorting because the way they are selected is very different from the way internal faculty is selected, and it varies substantially depending on specific departments and academic years. I therefore compare internal faculty only, depending on their year of entry and year of exit. This provides some suggestive evidence on the way selection of internal teachers changed over time.

I evaluate how the probability of exiting or entering the university for better researchers or teachers changed after the policy.

For sorting out, I estimate the following equations, for the population of teachers employed by Bocconi between 2001 and 2005 and between 2006 and 2011, respectively:

$$exit_p = \alpha_1^j \widehat{\theta_p^j} + \alpha_2^j X_p + \delta_e^j + u_p^j \tag{9}$$

where: $j = r, t$ refers to research and teaching, respectively; $exit_p$ is a dummy equal to one if teacher

[22] I only observe whether they were teaching undergraduate compulsory courses between 2001 and 2011, but not their exact year of entry/exit.

p left Bocconi in the considered years; δ_e^j are year of entry fixed effects; X_p are time-invariant professors' characteristics (age of entry, gender) and u_p^j is an error term and $\widehat{\theta_p^j}$ is defined above. I only include teachers leaving Bocconi for reasons different from retirement and who entered after 2000, to make the samples more comparable. Standard errors are bootstrapped.

Symmetrically, I obtain the effects on sorting in of teachers, by estimating the following equation, for the years 2001-2005 and 2006-2011:

$$entry_p = \beta_1^j \widehat{\theta_p^j} + \beta_2^j X_p + f^j(e_p) + \omega_p^j \tag{10}$$

where: $j = r, t$ refers to research and teaching, respectively; $entry_p$ is a dummy equal to one if teacher p joined Bocconi in the considered years; $f^j(e_p)$ is a linear and squared trend for year of entry; X_p are time-invariant professors' characteristics (age of entry, gender) and ω_p^j is an error term. To make the two groups of teachers more comparable, I estimate equation 10 only for teachers who entered after 2000. Standard errors are bootstrapped.

By comparing the coefficients of the effect of professors' skills on the probability of exiting/entering the university before and after 2006, I evaluate the changes in the selection pattern of teachers.

5 Results

5.1 Effects on research performance

The sign of the effect on research performance is expected to be positive, and stronger for young professors not tenured yet, since they are affected both by the changes in monetary incentives and by the changes in the promotion strategy.

Table 3 shows some descriptive statistics for the number and the quality of publications and working papers for internal and external teachers before and after 2006. The first panel analyses the total number of publications (books or journal articles) of professor p in year t, as collected from the Web of Science database. The second panel looks at the number of publications, weighted by the importance they have in terms of Bocconi's new incentive regime. This allows me to evaluate quality as well as quantity of research. Finally, the third panel evaluates the effect on the number of working papers obtained from Google Scholar.[23] The first column reports the mean and the standard deviation of publications for internal and external teachers. The second and the third columns break down the number of publications for the period before and after 2006. Finally, the number in the bottom-right corner represents the simple difference in difference, without any control. Standard errors, clustered at the teacher level, are reported in parenthesis. The Table shows that the number and the quality of publications increased after 2006 and they increased much more for internal professors than for external professors.

While the previous table shows that internal and external professors are rather different in terms of level of research activity, the validity of my identification strategy relies on the presence of parallel trends before 2006. Figure 2 shows my main results: it displays the evolution of the difference in the average number of publications between internal and external faculty over time (the omitted category being the year 2001). It therefore gives information both on the presence of parallel trends before 2006 - by evaluating whether the line is close to zero till 2005- and on the evolution over time of the effect of the new incentive scheme.[24] The dotted lines refer to the 10% confidence interval boundaries. While the difference is rather stable before 2005 (implying that the trend was parallel before 2006), it gets larger and positive after the introduction of research incentives (implying that there is a positive effect on research for internal professors). Moreover, given the long time needed to publish papers in most disciplines, after 2006 there is a clear change in trends but there is not a sharp jump.

Table 4 shows results from equation 3, using the three dependent variables described above. Columns (1) and (2) report estimates without teacher fixed effects. The effect is positive and significant in all three panels. Once I include teacher fixed effects (columns (3) and (4)), and I therefore control for changes in the composition of teachers, the effect is still positive and significant.

[23] Accessed in july 2011.

[24] This graph plots the coefficient γ_s of the following equation (where $\theta_s = 1$ if $t = s$):

$$pub_{py} = \theta_y + \theta_p + \Sigma_{s=2001}^{2011} \gamma_s (internal_p * \theta_s) + \gamma_4 Q_{py} + \eta_{py}$$

The dots therefore refer to the difference in performance between internal and external teachers between 2001 (the omitted year) and all following years.

After the introduction of research incentives, the number of publications increased by 0.14 (9% of a standard deviation) for internal faculty and the index used by Bocconi to evaluate teachers increased by 0.13 (6% of a standard deviation). The number of working papers of internal professors is also 0.15 (6% of a standard deviation) higher than it would have been otherwise. Moreover, while columns (1) and (3) look at the aggregate effect, columns (2) and (4) separately evaluate the effect for assistant and associate professors (which I call junior faculty) and full professors. The aggregate effect is mostly driven by junior faculty, as their career concerns are stronger.[25]
Finally columns (5) and (6) report results from estimating equation 3, using as dependent variable the square root of the number of publications. This is to tackle simultaneously the presence of possible outliers and of a lot of zeros.[26]

5.2 Effects on research spillovers

This section explores whether the presence of stronger research incentives for internal teachers has generated positive spillover effects. It is important to evaluate the presence and the relative magnitude of spillover effects for internal and external faculty, because they may amplify or shrink the effect of interest. The presence of spillover effects may change the way we interpret our results. If spillovers are present, but stronger for internal faculty, then they represent a mechanism that amplifies our effect of interest: the new incentive scheme did not only induce internal faculty members to put more effort into research but also to benefit from the interaction with better motivated researchers. If instead spillovers are larger for external faculty members, the opposite is true. When spillovers are symmetric, because they affect internal and external teachers in the same way, than the coefficient we estimate on the main regression (equation 3) represents the effect of research incentives, net of peer effects.

We study interactions by looking at two outcomes: the share of articles co-authored with other members of Bocconi internal faculty and the share of articles co-authored with individuals hired by Bocconi in the two previous years, to evaluate whether the possible hiring of prominent scholars over

[25] I do not divide external faculty into junior and senior faculty because in many cases I do not have information on their position in their institution of origin. However in Section 6 I use age as a proxy for junior and senior faculty in my control group and results are very similar.

[26] Moreover I dropped the 5/1000 highest values for each dependent variable, since it is very likely that most outliers are generated by homonymity.

that time period has generated important peer effects. While co-authorship is only one example of all possible forms of academic interactions, it is likely the most important, as suggested in Waldinger [2012].

Columns 1 and 2 of Table 5 shows that, even if peer effects may be present, they affected internal and external teachers in the same way: both coefficients are not statistically different from zero.

Columns 3 replicates the results obtained in Table 4, but columns 4 and 5 split the publications between those co-authored with internal faculty members and all other publications (not co-authored or co-authored with other scholars). The Table shows that most of the increase comes from the second component.

Overall, the results suggest that, even if peer effects may be present, they affected external and internal teachers in the same way.

5.3 Effects on teaching performance

As shown in Section 2, the sign of the effect of stronger research incentives on teaching quality depends on whether teaching and research efforts are complements or substitutes in the professors' cost function (δ smaller or larger than 0 respectively). The effect moreover is expected to be stronger for junior professors, exposed both to the change in monetary incentives and to the change in the minimum number of publications required.

Table B.3 presents the results obtained from estimating equation 4: my first step regression, aimed at obtaing a measure of teaching performance. Exam grades are standardized to have mean 0 and standard deviation 1 within the same course-year.[27] The results show that being male, with a higher final high-school grade, Italian and from Milan is associated with higher university exam grades.

Table 6 reports some summary statistics of the estimated α_{pcy} for internal and external teachers, before and after 2006. While before 2006, the teaching performance of the two groups was very similar, after 2006 it improved much more for external teachers than for internal teachers. Again, the bottom-right corner reports the difference in difference effect, without any control.[28]

[27] Grades in Italy go from 18 (pass) to 31 (excellence).

[28] Notice that, because of some sampling error generated by the fact that $\alpha_{\hat{pcy}}$ are estimated, the reported standard deviation may be larger than the standard deviation of the true α_{pcy}.

Figure 3 shows the evolution over time of the different performance of external and internal teachers (panel a) and of external teachers and assistant professors (panel b). Again, this allows me to test for the presence of parallel trends before 2006.[29] The difference is rather stable before the academic year 2005/2006 (named 2006 in the graph), implying that the trend was parallel before 2006. Right after the adoption of the new incentive regime there is a drop in the quality of teaching for internal professors. In the following years, the performance is still slightly worse than before the reform, but better than in 2006. This may be because internal professors understood the consequences of their effort reallocation and partially readjusted their behaviours. Alternatively, they just started being more generous with their grading standards. Section 5.4 analyzes this aspect in more details.

Table 7 displays results from estimating equation 5. The first two columns show results without teacher fixed effects. Column (3) and (4) add teachers fixed effects and show that teaching quality of internal teachers is 0.04 (around 7% of a standard deviation) lower after the change in incentives than it would have been otherwise. This suggests that teaching and research efforts are substitutes in the professors' cost function. Again, the effect is stronger for young faculty members, also exposed to career concerns.

Table 8 reports results from the student level regression (equation 6). As expected, results are very similar. What differentiates columns (1) and (2) of Table 8 from columns (3) and (4) of Table 7 is the way observations are weighted and coefficients should be interpreted. Table 8 implicitly weights observations by the number of students in each class: the coefficients should be interpreted as effects on average students' performance. Table 7 weights observations by teachers and the coefficients should be interpreted as effects of average teachers' performance.

Columns (3) and (4) of Table 8 explore whether the main results of Table 7 mask some important heterogeneity at the student level. I estimate equation 8, interacting the main effect with a proxy for students' ability. In particular I use high school final grade as proxy.[30] My omitted category are

[29]This is obtained by plotting the coefficients γ_s obtained from the following equation (where $\delta_s = 1$ if $t = s$):

$$\alpha_{pcy} = \delta_p + \delta_{cy} + \Sigma_{s=2001}^{2011}\gamma_s(internal_p * \delta_s) + \gamma_2 Q_{py} + \gamma_3 Z_{pcy} + \epsilon_{pcy}$$

Year 2001 (and the interaction between 2001 and internal) is omitted. The dotted lines refer to the 10% confidence interval bands.

[30]I divide it into 3 categories: (i) high ability (omitted)= those students whose final high school grade was between

high ability students. Results show that the negative effect is mostly borne by low ability students. This result suggests that, in order to reduce the overall negative effect of stronger research incentives on teaching performance, there is room for policies aimed at matching professors more under pressure to students who are less damaged.

5.4 Robustness checks for the effect on teaching

Table 9 presents a set of robustness checks for the estimation of the teaching equation. First, I estimate equation 5 excluding the academic years 2008/2009, 2009/2010 and 2010/2011. Starting from 2008/2009, internal faculty was exposed not only to research incentives, but also to teaching performance monetary awards. In particular, Bocconi University created a commission in charge of awarding a premium of 20,000 euros for the best 20 teachers who voluntarily apply. Decisions are based on students' evaluations. This new policy may attenuate the effect of research incentives (Holmstrom and Milgrom [1991, 1994]). Column (1) of Table 9 shows that results are almost unchanged. Second, in Column (2) I include lecturers in my sample and I estimate a different treatment effect for lecturers. The effect on internal professors remains similar. The effect on lecturers, even if not significant because of the small number of observations, is negative. Column (3) includes endogenous switches from internal to external status after the policy: it uses the contemporaneous status, not the status before 2006 as in Table 7, to define internal status. The control group includes in this case also, for instance, professors who switched from being internal to being external as a consequence of the policy. The coefficient is still negative and significant, but the magnitude is smaller. This means that Bocconi promotions from external to internal where positively correlated with teaching quality. In column (4) I weight my regression by the number of hours taught by each professor in each class. The results are again very similar to what found in Table 7.

I now discuss three possible confounding factors, that may undermine my identification strategy. The first is that students might not comply with the random class assignment and they might endogenously decide to attend classes with different lecturers. For example, they may match to the best professors, or attend classes with their closest friends. Unfortunately, I do not have any

1 and 0.9; middle ability = between 0.8 and 0.9 and low ability: below 0.8.

direct information on these unofficial switches of classes.[31] Braga et al. [2014] analyze whether the direction of class switches at Bocconi University is correlated with professors' ability between 1998 and 1999. They use data on students' answers to an item in the student evaluation forms asking them about the level of congestion in their classroom. They estimate the degree of class switches as the difference in congestion levels between the most congested and the least congested classes for each course. They find that, overall, course switching is not related to teacher effectiveness in any direction. Therefore, if the process of class switching is unrelated to teachers or students quality, then it will just affect the precision of my estimated class effects. Moreover, if the process is constant over time, the effect will go away with professors' fixed effects. Finally, even if course switching does affect my results, it would probably bias them against finding a negative effect on teaching performance. It is likely that students, if anything, will react by attending classes with the best teachers, who after the change of incentives will more likely be external faculty members. This would reduce the negative effect of the incentive policy on teaching.

Another concern is that teachers may change the way they grade students' exams as an effect of observing worse performances of their students. In this case the observed change in teaching quality may actually be confused with a change in grading standards. There is not a common rule on how exams are graded at Bocconi: in some cases exam papers are randomly allocated to be graded to one of the course teachers independently of the class they were assigned to, in some other cases each professor is in charge of grading his own group. I do not have information on how exam papers are actually graded in each course. However, in the fist place, if anything, I expect internal teachers to start being more lenient towards their students, therefore I expect this type of bias to go against finding a negative effect on teaching performance. Moreover Table 10 addresses this point by looking at subjects where grades are more difficult to be manipulated. Columns (1) and (2) look at the effect on teaching quality for exams that are more objectively-graded, such as math, statistics or quantitative finance. Results show that, even if the effect is slightly smaller and less precise for this types of courses, it remains negative.

[31] It would have been in principle possible to grasp the size of classes reallocation by using students' evaluations, exploiting the information for whether the number of answers is larger or smaller than the official class size. However, Bocconi decided to hand in evaluation forms to a sub sample of professors only exactly in the year 2005 and 2006, making it impossible to look at students' evaluations for the period when the policy took place.

Another concern is that internal faculty may have managed to reduce its teaching loads and to avoid some of its previous teaching responsibilities, under the new regime. In columns (3) and (4) of Table 10, I check whether the new incentive scheme implied a change in internal teachers' teaching duties. I estimate equation 5 using as dependent variables a dummy equal to one if professor p was the course coordinator in year t and the number of teaching hours taught by professor p in year t, respectively. Results show that there is no significant change in the type of teaching loads and duties before and after the change in the incentive regime. This suggests that the change in teaching quality was not driven by other, simultaneously related, changes in how teaching was organized and distributed.

Finally, in Table 11, I check whether my results may be driven by a recommendation letter sent by Bocconi University in 2006 to the entire teaching faculty, asking for higher homogeneity of grades across classes. This may affect my analysis, if internal and external teachers responded to this request differently. Table 11 displays the standard deviation of average class grades across classes belonging to the same course and degree program, by academic year and by whether the teacher was internal or external. The variability of grades between classes did not decrease right after 2006, as a consequence of such recommendation, and there was no differential response between internal and external teachers.

5.5 Teaching and research skills

Understanding the sign of the correlation between α_r and α_t, as defined in Section 2, is crucial both to have a full picture of potential selection effects and to understand the plausible cost of separating careers of teachers and researcher in university.

Figure 4 and Table 13 correlate the two sets of fixed effects as estimated from equations 7 and 8 and they show that teaching and research ability are strongly positively correlated: good researchers are also good teachers. This is an important result that has not been estimated before. Columns (1) and (3) include all teachers in my sample. Columns (2) and (4) try to address the fact that teacher fixed effects represent noisy measures of the true teaching and research abilities and sampling error may bias the coefficients of columns (1) and (3). I exploit the fact that sampling error decreases substantially if the analysis is performed on a subsample of teachers with a large number of observations. I therefore estimate the correlation, including only teachers for which I can estimate the fixed effects with more that 5 observations.[32] Results are very similar but, as expected, after the correction the coefficients are larger, because not affected anymore by the attenuation bias.

The positive correlation between research and teaching skills and the large standard deviation of the teachers fixed effects have important policy implications. First, comparing the standard deviation of the fixed effects plotted in Figure 4, which quantify the variation in the time-invariant part of teaching quality (in teaching ability), with the coefficients obtained in Tables 4 and 7, it is clear that sorting effects may potentially have much larger and substantial consequences on the overall research and teaching productivity than substitution effects. Keeping the composition of teachers constant (i.e. looking at the intensive margin), the reform of the incentive structure improved research productivity by around 9% of a standard deviation and decreased teaching quality by 7% of a standard deviation. Instead, when we allow the composition of teachers to change (i.e. we look at the extensive margin) and we incorporate the fact that universities may, as a consequence, attract (push away) the very best (worst) researchers, the average productivity may potentially increase by much more. Therefore, even if Section 5.3 showed that, at the margin, pushing university professors to focus more on research may induce them to crowd out time from preparing teaching classes and may worsen their teaching performance, Figure 4 shows that selection effects may potentially be much more effective.

Second, the fact that teaching and research ability are positively correlated entails that if universities are able to attract/maintain good researchers, they will also, indirectly, improve teaching quality. A very popular proposal to solve the trade-off between teaching and research, is to increase specialization of faculty members. This would entail, for example, the creation of two groups of professors, one more research-oriented and one more teaching-oriented. Figure 4 and Table 13 show that these proposals should take into consideration that good researchers are also good teachers and the potential benefit of separating careers may be minimal, given that the trade-off on the intensive margin (generated by forcing good researchers to teach some class instead of researching)

[32] Notice that I always estimate the research fixed effect with 10 (yearly) observations. For the teaching fixed effects, instead, the number of teacher-specific observations used depends on the number of time I observe teacher p teaching undergraduate compulsory courses.

is less sizable than the trade-off on the extensive margin (generated by excluding good researchers from teaching, for instance).

5.6 Effects on the composition of faculty members

The first way I analyse selection effects is by evaluating the difference between the OLS and the fixed effect estimates in Tables 4 and 7. OLS estimates are always larger than fixed effects estimates, suggesting that the policy induced some positive selection effects.

I also analyse sorting in and out separately using direct estimates of teachers' underlying skills, obtained through equations 7 and 8.

Table 14 (15) shows how the probability of entering or exiting Bocconi university for good researchers (teachers) has changed after 2006. I particular, we estimate equations 9 (10) and we display how the coefficient on the internal professors' research (teaching) ability changes for the period before and after 2006. The first four columns evaluate the effects on the probability of exiting Bocconi university (for reasons different from retirement) and use as main independent variable the proxy for research (teaching) skills obtained from Section 5.5. I use professor fixed effects estimated on the entire sample (columns (1) and (2)) and on the sample pre 2006 only (columns (3) and (4)). The results, in line with the predictions of Section 2, show some evidence that the change in incentives induced worse teachers to leave.[33] Columns (5) and (6), look instead at the probability of entering Bocconi university. Here the results are much weaker. The reason why I don't find any positive sorting-in effect may be due to the fact that it takes time to publish papers and my proxy for teaching and research skills may be weaker for very young scholars.

5.7 Heterogeneity by teachers' ability

This Section analyses how the effect on teaching and research performances changes with respect to teachers' ability. Section 2 shows that the bulk of the effort reallocation should be concentrated on low ability researchers, while the effect on research should be concentrated on very low ability (because of fair of being fired) or very high ability (because they benefit more from any unit of

[33] Even if the coefficient of column (4) is not significant, the sign and the magnitude of the coefficient confirm the higher probability of exiting for low skilled researchers.

effort in research) researchers.

Column 1 of Table 12 shows that the negative effect on teaching activity is stronger for low ability researchers than for higher ability ones. The Table displays the coefficients of the $internal_p * post2006_t$ dummy of equation 5 interacted with research ability tertiles.[34]

For what concerns the heterogeneity of the effect on research performance, columns 2 and 3 of Table 12 show that the positive effect is driven by low and middle ability researchers. The difference is more evident in column 3, that looks at the effect on the number of working papers.

6 Alternative control groups

One possible concern of using external teachers as control group is that these teachers may react to the policy as well if, for instance, their final objective is to be hired by Bocconi. This would spoil my identification strategy because it implies that the effect of the policy would spill over my control group. Moreover, one may think that external teachers are a natural control group for evaluating the effect on teaching performance but may not be as good as a control group for research activity, because they may have very different research productivity and may be on very different trends in any case. To tackle these issues I propose two alternative control groups.

The first one refers to the analysis on research. In Table 16, I use all professors belonging to Bologna University faculty in 2005 as alternative control group. Bologna University is another Italian University, whose department of management and economics is quite similar to Bocconi University in terms of quality of the economics/management department. Bologna university economics and management department is indeed ranked as the best[35] department among Italian public institutions.[36] Again, I obtain data on their publications from the Web of Science website and data on the faculty composition in 2005 from the website of the Italian Ministry of Education.[37]

The second alternative control group are professors who became tenured before the policy.

[34]Tertiles are calculated using θ_p^r of equation 8, and are estimated only for the years before the change in the incentive regime.

[35]Or one of the top three departments, depending on the type of ranking in 2011.

[36]The average yearly number of publications of Bologna junior (senior) professors is 0.2 (0.3), and the average number of publications of Bocconi junior (senior) professors is 0.4 (0.5), based on faculty composition in 2005. Source: Web of Science, 2001-2011.

[37]www.miur.it

Given that the change in the incentive structure acts mainly in terms of promotions and tenure decisions, full professors should only be marginally affected. Since they are already hired on a full time base by Bocconi, they should not react to changes in hiring/promotion strategies. If we assume that the effect of monetary incentives is the same on full and junior professors, than what I estimate using full professors as control group is the effect of the change in career requirements only. However, for publications it is very likely that trends for tenured and non tenured professors are different, since tenure decisions are based research productivity or potential productivity. I will therefore use this alternative control group only for the analysis on teaching.

Moreover, both for the analysis on teaching and on research, I estimate my difference-in-difference models separately on two sub samples of professors with similar age. In particular I split both the sample of internal and external teachers between those older than 43 (the median age) and those younger than 41. This allows me to use as control group for young researchers, young external researchers and viceversa for senior faculty.

Table 16 reports results for research activity. In columns (1) and (2) I run equation 4 on the subsamples of teachers younger and older than 43, respectively. As for the main results, the effect is larger for junior professors. Columns (3), (4) and (5) use, instead, Bologna faculty members as control group. Column (3) looks at the aggregate effect, column (4) compare junior professors at Bocconi with junior professors at Bologna University and column (5) compares full professors in the two universities. The effect is remarkably similar to my baseline estimates. The introduction of incentives led to an increase in the number of publication of 0.17 for Bocconi faculty members, very similar to the effect found in Table 4. The increase is stronger for young faculty members.

Table 17 reports, instead, results using the proposed alternative control groups for teaching quality. Columns (1) and (2) split the sample by age. The effect is similar to what found in my baseline estimates and is more negative for junior professors. Columns (3) and (4) use full professors as alternative control group. Columns (3) does not include teacher fixed effects. Column (4) shows results including teachers fixed effects. Again, results are remarkably similar to what found in Table 7. The introduction of research incentives worsened teaching performance by 0.04, about 7% of a standard deviation. I can't use Bologna faculty members as control group for the analysis on teaching, because information on teaching performance of Bologna faculty members is not publicly available.

Figure 5 checks the presence of parallel trends for both alternative control groups, in the same way I checked for parallel trends in sections 5, by plotting the evolution over time of the difference in the performances of internal and external teachers. The figure shows that trends are parallel for both alternative control groups.

7 Conclusions

This paper exploits a natural experiment to evaluate the magnitude of possible behavioural responses predicted by models of incentives in a multi-task environment.

I use administrative data from Bocconi University to analyse the faculty reaction to a sharp increase in research incentives. The heterogeneity in the type of teaching faculty's contracts provides me with a suitable control group for my Difference-in-Difference estimation. The randomization of teachers to students within the same course, in a context where the syllabus and the exams are fixed, allows me to build a credible measure of teaching performance. In particular, the specific Bocconi setting allows me to overcome two of the reasons why analyses of teachers' effectiveness are rarely done at the post secondary level: the lack of standardized tests and the endogeneity in students selection of courses (and professors).

I find evidence that the introduction of incentives on one task only (research) affects the allocation of effort across all tasks. My results show that professors' teaching performance gets worse while their research performance significantly improves. The effect is stronger for young faculty members, also exposed to career concerns, in line with the predictions of Holmstrom and Milgrom [1991, 1994]. The number of working papers and published papers of internal Bocconi faculty increases after the introduction of incentives on research, in line with Prendergast [1999], Lazear [2000], Checchi et al. [2014], for example. The effect on quantity of publications moreover does not go against the quality of publications. This may be due to the way research incentives are structured by Bocconi. On the other hand, teaching performance of faculty members more exposed to research incentives is 7% of a standard deviation worse after the change in the incentive regime. The effect is nonproportionally borne by lower ability students. My estimates suggest that encouraging one more paper has an implicit cost of 0.3 standard deviation on teaching quality. Finally, I find evidence of positive selection effects: after the change in incentives, lower quality researchers are more likely to leave Bocconi and, since teaching and research ability are positively correlated, also worse teachers are induced to leave.

My results suggest that it is beneficial to evaluate new policies not in isolation but as part of a coherent incentive system. I believe this paper delivers three policy-relevant messages. First, the effects are heterogeneous: the negative effect on teaching is mostly borne by low ability students and it is mostly generated by lower skilled researchers, leaving room for systems of allocation of tasks and courses to teachers that minimize possible distortions. Second, selection effects may countervail effort substitution effects: I find that, while at the margin there is a trade-off between teaching and research, i.e. incumbent professors are induced to decrease teaching effort, universities are also able to keep the best researchers under the new incentive regime and, since good researchers are also good teachers, overall teaching quality may improve as well. Finally, I provide the first evidence on the correlation between research and teaching ability. This has important implications for the design of professors' incentives and hiring schemes. Policies aimed at increasing teachers' specialization that propose to dedicate part of the faculty exclusively to teaching and part of it exclusively to research, should take into consideration that there is substantial overlap between good researchers and good teachers.

References

Daniel Aaronson, Lisa Barrow, and William Sander. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1):95–135, 2007.

Joshua D. Angrist and Victor Lavy. Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement. *Quarterly Journal of Economics*, 114(2):533–575, 1999.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Incentives for managers and inequality among workers: evidence from a firm-level experiment. *Quarterly Journal of Economics*, 122(2):729–773, 2007.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, 2009.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social incentives in the workplace. *Review of Economic Studies*, 77(2):417–458, 2010.

Roland Benabou and Jean Tirole. Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70(3):489–520, 2003.

Eric Bettinger and Long Bridget Terry. Does Cheaper Mean Better? The Impact of using Adjunt Instructors on Student Outcomes. *Review of Economics and Statistics*, 92(3):598–613, 2010.

Ernest L. Boyer. Scholarship Reconsidered Priorities of the Professorate. Technical report, Princeton, N.J. : Carnegie Foundation for the Advancement of Teaching, 1990.

Michela Braga, Marco Paccagnella, and Michele Pellizzari. Evaluating Students Evaluations of Professors. *Economics of Education Review*, 41:71–88, 2014.

Michela Braga, Marco Paccagnella, and Michele Pellizzari. The Academic and Labor Market Returns of University Professors. *Journal of Labor Economics*, 34(3):781–822, 2016.

James A. Brickley and Jerold L. Zimmerman. Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate Finance*, 7(4):367–396, 2001.

Alexander Brüggen and Frank Moers. The role of financial incentives and social incentives in multi-task settings. *Journal of Management Accounting Research*, 19(1):25–50, 2007.

Scott E. Carrell and James E. West. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy*, 118(3):409–432, 2010.

Daniele Checchi, Gianni De Fraja, and Stefano Verzillo. Publish or Perish? Incentives and Careers in Italian Academia. IZA Discussion Papers 8345, Institute for the Study of Labor, 2014.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9):2593–2632, 2014.

Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5):1739–74, 2011.

Etienne Dumont, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of Health Economics*, 27(6):1436–1450, 2008.

Glenn Ellison. Evolving Standards for Academic Publishing: A q-r Theory. *Journal of Political Economy*, 110(5):994–1034, 2002a.

Glenn Ellison. The Slowdown of the Economics Publishing Process. *Journal of Political Economy*, 110(5):947–993, 2002b.

Ernst Fehr and Urs Fischbacher. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*, 112(478):C1–C33.

Susan Feng Lu. Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes. *Journal of Economics Management Strategy*, 21(3):673–705, 2012.

David N. Figlio and Lawrence W. Kenny. Individual Teacher Incentives and Student Performance. *Journal of Public Economics*, 91(5-6):901–914, 2007.

David N. Figlio, Morton O. Schapiro, and Kevin B. Soter. Are Tenure Track Professors Better Teachers? *Review of Economics and Statistics*, 97(4):715–724, 2015.

Roland G. Fryer and Richard T. Holden. Multitasking, Dynamic Complementarities, and Incentives: A Cautionary Tale. Technical Report 17752, NBER working paper, 2012.

Michael Gibbs, Kenneth Merchant, Wim Van der Stede, and Mark Vargus. Performance Measure Properties and Incentives. 2004.

Giacomo De Giorgi, Michele Pellizzari, and William Gui Woolston. Class Size and Class Heterogeneity. *Journal of European Economic Association*, 10(4):795–830, 2012.

Bengt Holmstrom and Paul Milgrom. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization*, 7(0):24–52, 1991.

Bengt Holmstrom and Paul Milgrom. The Firm as an Incentive System. *American Economic Review*, 84(4):972–91, 1994.

Fuhai Hong, Tanjim Hossain, John A. List, and Migiwa Tanaka. Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories. Working paper, National Bureau of Economic Research, 2013.

Jun Ishii and Steven G Rivkin. Impediments to the Estimation of Teacher Value Added. *Education Finance and Policy*, 4(4):520–536, 2009.

Brian A. Jacob. Accountability, Incentives and Behavior: the Impact of High-stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5-6):761–796, 2005.

Carlos Munoz Johnson, Ryan and David H. Reiley. The War for the Fare': How Driver Compensation Affects Bus System Performance. *Economic Inquiry*, forthcoming.

Thomas J. Kane and Douglas O. Staiger. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. NBER Working Papers 14607, National Bureau of Economic Research, Inc, 2008.

Edward P. Lazear. Performance Pay and Productivity. *American Economic Review*, 90(5):1346–1361, 2000.

Daniel McCaffrey and J.R. Lockwood. Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics*, 29, 2004.

Abigail Payne and Joanne Roberts. Government Oversight of Public Universities: Are Centralized Performance Schemes Related to Increased Quantity or Quality? *Review of Economics and Statistics*, 92(1):207–212, 2010.

Canice Prendergast. The Provision of Incentives in Firms. *Journal of Economic Literature*, 37(1): 7–63, 1999.

Canice Prendergast. Intrinsic Motivation and Incentives. *American Economic Review*, 98(2):201–205, 2008.

Steven G Rivkin, Eric A Hanushek, and John F Kain. Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458, 2005.

Jonah E Rockoff. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, 94(2):247–252, 2004.

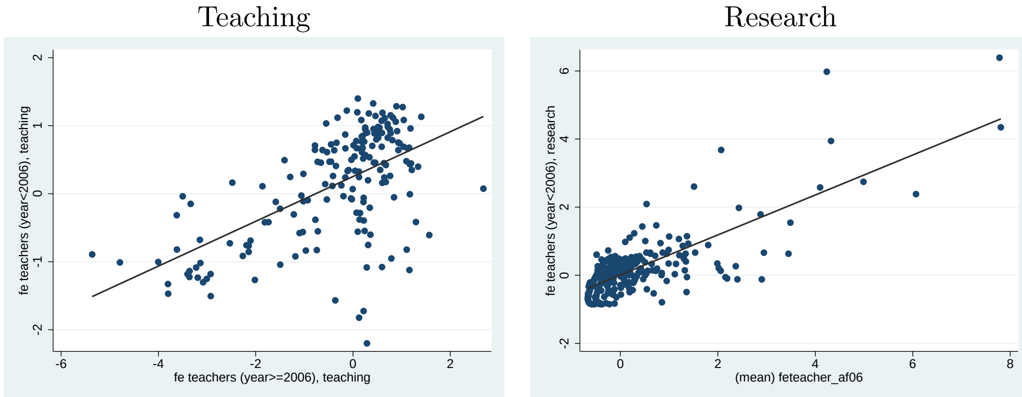
Jesse Rothstein. Student Sorting and Bias in Value Added Estimation: Selection on Observables and Unobservables. *Education Finance and Policy*, 4(4):537–571, 2009.

Jesse Rothstein. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Quarterly Journal of Economics*, 125(1):175–214, 2010.

Margaret E. Slade. Multitask Agency and Contract Choice: An Empirical Exploration. *International Economic Review*, 37(2):465–486, 1996.

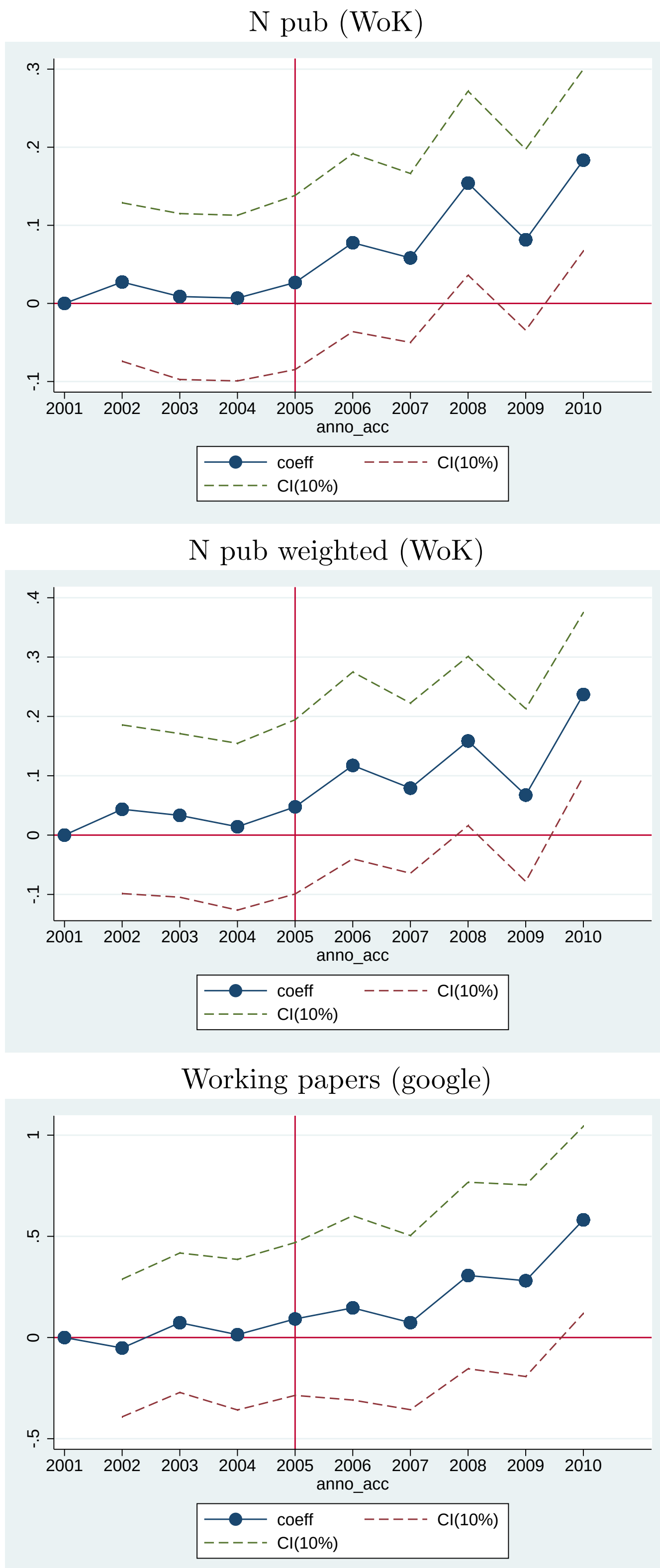
Fabian Waldinger. Peer Effects in Science – Evidence from the Dismissal of Scientists in Nazi Germany. *Review of Economic Studies*, 79(2):838–861, 2012.

Figure 1: Robustness of teachers fixed effects



Source: Bocconi registers and Web of Science and Google Scholar. On the x axis=fixed effects based on the post 2006 sample; on the y axis= fixed effects based on the pre 2006 sample.

Figure 2: Research difference in difference graphs: parallel trends

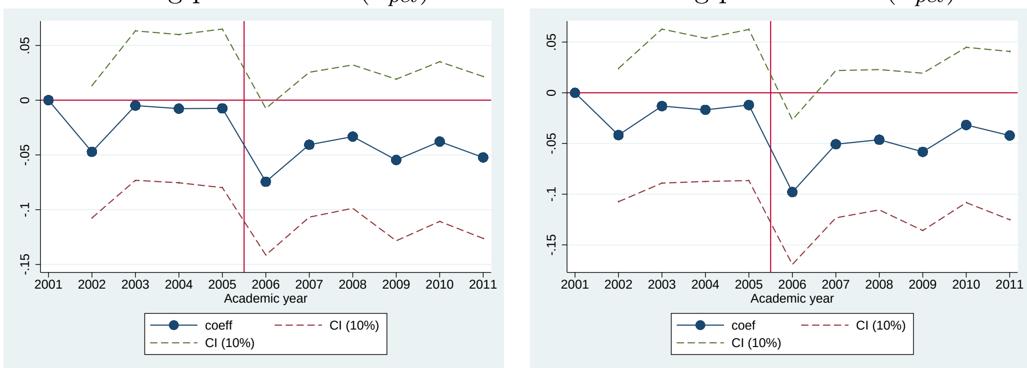


Source: Web of Science and Google Scholar. The solid line displays the coefficient of the interaction between the year dummies and the internal professor (in 2005) dummy (γ_s); the dashed lines represent the 10% confidence interval where standard errors are clustered by teacher.

Figure 3: Teaching difference in difference graphs: parallel trends

Internal professors
teaching performance (α_{pct})

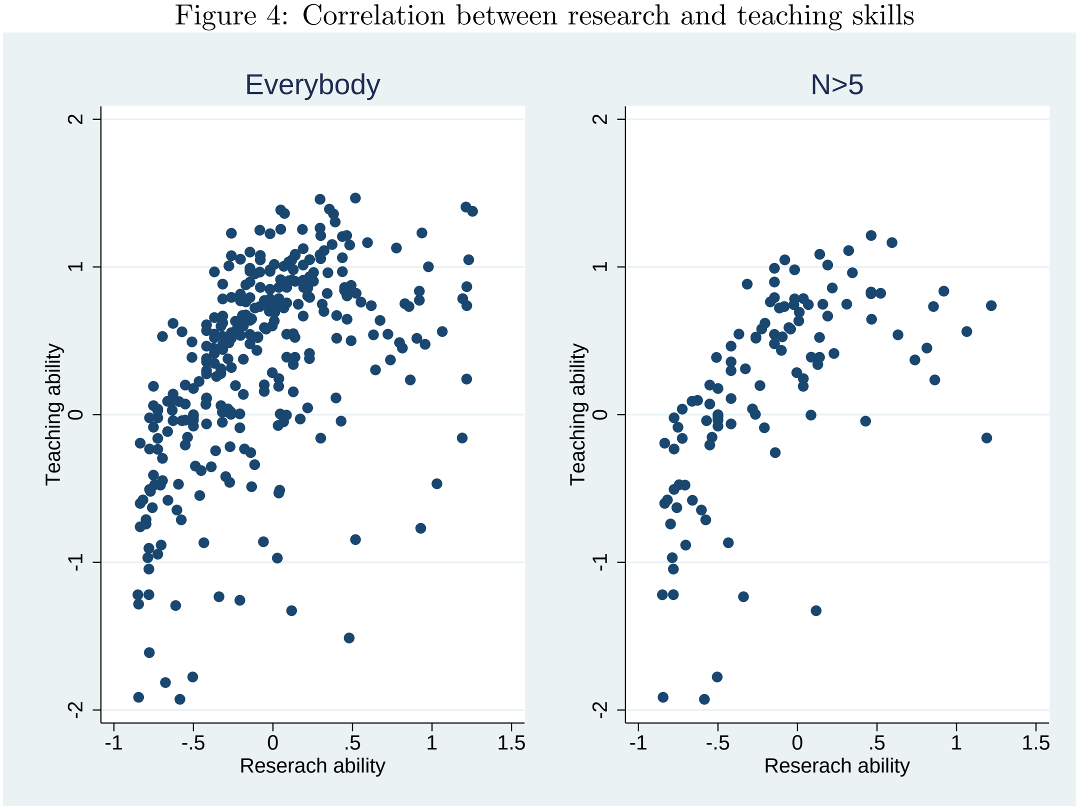
Assistant professors
teaching performance (α_{pct})



a b

Source: Bocconi student and teacher archives. The solid line displays the coefficient of the interaction between the year dummies and the internal professor (in 2005) dummy (γ_s); the dashed lines represent the 10% confidence interval where standard errors are clustered by teacher.

Figure 4: Correlation between research and teaching skills

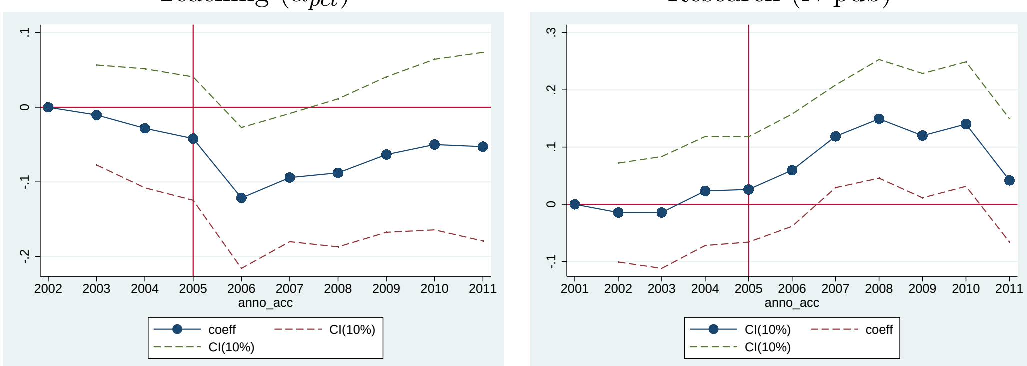


Source: Bocconi registers and Web of Science and Google Scholar. Panel 1 plots $\widehat{\theta}_p^t$ and $\widehat{\theta}_p^r$ estimated on the entire sample of teachers. Panel 2 uses only the $\widehat{\theta}_p^t$ estimated on at least 5 observations.

Figure 5: Alternative identification strategies graphs: difference in difference graphs - parallel trends

control: full prof
Teaching (α_{pct})

control: Bologna prof
Research (N pub)



Source: Bocconi registers and Web of Science and Google Scholar. The solid line displays the coefficient of the interaction between the year dummies and the internal professor (in 2005) dummy (γ_s); the dashed lines represent the 10% confidence interval where standard errors are clustered by teacher.

Tables

Table 1: Descriptive statistics - Students

Variable	Mean	Std. Dev.	Min.	Max.
	[1]	[2]	[3]	[4]
1=female	0.469	0.499	0	1
year birth	1985	3.249	1954	1993
1=italian	0.973	0.163	0	1
1=from Milan	0.246	0.431	0	1
High School grade	0.899	0.103	0.6	1
Av. exam grades at Booconi	25.532	3.532	18	31
N		501189		

The sample consists of students taking compulsory undergraduate courses between 2001-2011. High School grade normalized to be between 0 and 1 (pass if >= 0.6) for all countries.

Table 2: Descriptive statistics - Teachers

	Internal	External	Diff
Teachers' descriptives			
N teaching hours per class	38.91	33.91	5.47***
	(16.60)	(17.44)	(1.34)
Age	43.18	41.29	1.89**
	(9.45)	(7.80)	(0.77)
% female	32.27	34.25	-0.20
	(0.47)	(0.47)	(0.045)
Teachers' Department			
Accounting	14.8 %	20.8%	
Math/Stat	13.3%	24.6%	
Economics	20.2%	13.8%	
Finance	16.7%	7.4%	
Management	39.0%	33.5%	
Teachers' Contracts			
% Assistant prof	50.04%		
% Associate prof	10.45%		
% Full prof	12.65%		
% Non academic		9.61%	
% Other univ prof		9.11%	
% Lecturers	7.76%		

Standard deviation (columns 1 and 2) and standard errors (column 3) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 3: Summary statistics-Research

	Overall	Post 2006	Pre 2006	Diff
		N publications		
Internal[a]	0.539	0.680	0.302	0.381***
sd	1.561			(0.061)
External	0.416	0.481	0.264	0.239***
sd	1.318			(0.042)
Diff	0.199	0.199**	0.037***	0.143**
	(0.199)	(0.100)	(0.071)	(0.074)
		N publications (Bocconi index[b])		
Internal[a]	0.814	0.927	0.625	0.336***
sd	2.328			(0.084)
External	0.575	0.634	0.437	0.251***
sd	1.944			(0.080)
Diff	0.293	0.293**	0.187*	0.085
		N working papers (Google Scholar)		
Internal[a]	1.506	1.692	1.193	0.526***
sd	2.583			(0.126)
External	1.052	1.159	0.809	0.343***
sd	2.278			(0.105)
Diff	0.533	0.533**	0.385***	0.182*
	(0.533)	(0.172)	(0.191)	(0.164)

Source: panel 1 and 2 Web of Science; panel 3: Google Scholar. Standard deviation (column 1) and standard errors (column 4, last row) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] This refers to the status in 2005.

[b] Publications are weighted in the same way Bocconi University assigns monetary incentives. I give weight=15 if articles are in journals considered by Bocconi as belonging to band "A+", weight=7 if journals are considered as belonging to band "A", weight=3 if belonging to band "B" and weight=1 if not belonging to any band. The index is computed as $\Sigma_i(weight_i * pub_i)/Nauthors_i$ where i is a publication published by professor p in year t.

Table 4: Effect on research performance

Dep. var:	N Pub				N pub$^{(1/2)}$	
	[1]	[2]	[3]	[4]	[5]	[6]
Internal^a*post2006	0.206**		0.142**		0.081**	
	(0.100)		(0.070)		(0.034)	
Junior pr^a * post2006		0.224*		0.157*		0.087**
		(0.114)		(0.091)		(0.039)
Full pr^a* post2006		0.137		0.099		0.065
		(0.186)		(0.123)		(0.050)
N	5230	5230	5230	5230	5230	5230
Dep. var:	N Pub (weighted by Bocconi)^b				N pub w$^{(1/2)}$	
Internal^a*post2006	0.315***		0.130		0.082**	
	(0.103)		(0.098)		(0.041)	
Junior pr^a * post2006		0.266**		0.154		0.094**
		(0.110)		(0.109)		(0.047)
Full pr^a* post2006		0.496**		0.064		0.050
		(0.231)		(0.174)		(0.060)
N	5209	5209	5209	5209	5209	5209
Dep. var:	N working papers (Google Scholar)				N wp$^{(1/2)}$	
Internal^a*post2006	0.711***		0.148		0.091*	
	(0.166)		(0.139)		(0.052)	
Junior pr^a * post2006		0.492***		0.212*		0.120**
		(0.166)		(0.136)		(0.059)
Full pr^a* post2006		1.572***		-0.035		0.008
		(0.404)		(0.203)		(0.073)
N	5113	5113	5113	5113	5113	5113
Teacher fe	No	No	Yes	Yes	Yes	Yes

Additional controls: age, age squared, academic year fixed effects. Years between 2001 and 2011. Only professors included in the analysis on teaching. Junior professors are assistant and associate professors. The type of contract is defined according to the year 2005. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] This is the internal status in 2005

[b] Publications are weighted in the same way Bocconi University assigns monetary incentives. I give weight=15 if articles are in journals considered by Bocconi as belonging to band "A+", weight=7 if journals are considered as belonging to band "A", weight=3 if belonging to band "B" and weight=1 if not belonging to any band. The index is computed as $\Sigma_i(wight_i * pub_i)/Nauthors_i$ where i is a publication published by professor p in year t.

Table 5: Spillover effects on research performance

Dep. var:	% articles co-authored with		n articles published with		
	internal	new hires[a]	overall	internal[c]	others
	[1]	[2]	[3]	[4]	[5]
Internal[b]*post2006	-0.032	0.001	0.142**	0.049	0.093
	(0.023)	(0.023)	(0.070)	(0.033)	(0.075)
N	5230	4699	5230	5230	5230
Teacher fe	Yes	Yes	Yes	Yes	Yes

Additional controls: age, age squared, academic year fixed effects. Years between 2001 and 2011 (2003 and 2011 for new hires). Only professors included in the analysis on teaching. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Share of publications published with individuals who appeared in the sample as internal teachers the two years right before the cosidered year. Column 2 includes only years from 2003 onward, because otherwise for external teachers it is missing the information on the year of hiring.

[b] This is the internal status in 2005.

[c] N. of publications published with at least one internal professor; others are all other publications (with no coauthrs or with only external coauthors).

Table 6: Descriptives for teaching performance

		α_{pct}			
		Overall	Post 2006	Pre 2006	Diff
Internal[a]	*mean*	-0.020	0.146	-0.197	0.343***
	sd	(0.632)			(0.024)
External	*mean*	0.074	0.239	-0.192	0.431***
	sd	(0.645)			(0.026)
Diff			-0.093***	-0.005	-0.088***
			(0.033)	(0.015)	(0.036)

α_{pct} estimated from Table B.3, normalized to have mean 0. Standard deviation (column 1) or standard errors (column 4, last row) in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Based on their status in 2005.

Table 7: Step 2: regression at teacher level - students' grades

Dep. var:	α_{pct}			
	[1]	[2]	[3]	[4]
Internal[a]*post06	-0.011		-0.037**	
	(0.012)		(0.018)	
Junior pr[a]*post06		-0.014		-0.042**
		(0.013)		(0.020)
Full pr[a]*post06		-0.001		-0.023
		(0.016)		(0.022)
N	3889	3889	3889	3889
Teachers fe	No	No	Yes	Yes
Year*course*degree pr fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, class size, class average final high school grade. Junior professors are assistant and associate professors. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Based on status in 2005.

Table 8: Regression at the student level - heterogeneity by students' high school grades

	Dependent variable: stud grade (std)			
	[1]	[2]	[3]	[4]
int[a]*post06	-0.037***		0.002	
	(0.014)		(0.016)	
jun[a] pr*post06		-0.045***		-0.005
		(0.016)		(0.017)
full[a] pr*post06		-0.009		0.028
		(0.020)		(0.022)
int[a]*post06*mid ability stud			-0.079***	
			(0.014)	
int[a]*post06*low ability stud			-0.097***	
			(0.020)	
jun[a]*post06*mid ability stud				-0.077***
				(0.015)
jun[a]*post06*low ability stud				-0.100***
				(0.021)
full[a]*post06*mid ability stud				-0.086***
				(0.022)
full[a]*post06*low ability stud				-0.086**
				(0.036)
N	346628	346628	346628	346628
Teachers fe	Yes	Yes	Yes	Yes
Year*course*degree pr fe	Yes	Yes	Yes	Yes

Control set: teacher age, age squared, student gender, type of high school, whether Italian, whether from Milan. Ability based on final high school grade of students (normalized between 1 and 0, pass if >= 0.6): High ability (omitted)=between 1 and 0.9; middle ability = between 0.8 and 0.9; low ability0 below 0.8. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status as it was before 2006

[b] The number of observations is lower because Bocconi collected students evaluations in only a subsample of courses for the years 2004/2005 and 2005/2006.

Table 9: Robustness checks for the teaching regression

	no 09-10-11	also lecturers	include switches	weight by h. taught
	[1]	[2]	[3]	[4]
Internal[a]*post06	-0.037*	-0.034*		-0.035*
	(0.020)	(0.018)		(0.021)
Lecturer[a]*post06		-0.047		
		(0.042)		
Internal[b]*post06			-0.027*	
			(0.016)	
N	2848	4201	3889	3889
Teachers fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, teacher experience in Bocconi class size. Column (1) excludes the years when teaching incentives were also in place; column (2) includes lecturers and specifies a different treatment effect for lecturers; column (3) includes switchers and teachers fixed effects; column (4) weights professors by number of teaching hours. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status as it was before 2006

[b] Contemporaneous status

Table 10: Robustness checks for the teaching regression 2

Dep var:	Grading α_{pcy}		1=course coordin[a]	1=Num of taught h[b]
	[1]	[2]	[3]	[4]
int*post 06	-0.045**	-0.042**	0.025	0.671
	(0.020)	(0.020)	(0.037)	(1.084)
int*post 06*obj[c]	0.024			
	(0.047)			
int*post 06*math dep[d]		0.017		
		(0.046)		
N	3889	3889	3889	2989[e]
Teachers fe	Yes	Yes	Yes	Yes

Additional controls: age and age squared of teachers, dummies for teacher experience in Bocconi. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] 1=whether professor p in year t was the course coordinator

[b] Tot n of teaching hours in year t by professor p

[c] Objective if the name of the course includes the words *"math"*, *"stat"*, *"quantit"*

[d] Math if the teacher belongs to the math and statistics departments

[e] N of observations at the teacher-year level (if a teacher teaches more than one courses n of teaching hours are summed)

Table 11: Robustness checks for the teaching regression 3: average class grades

	sd av. class gr internal	sd av. class gr external
2001	0.390	0.434
2002	0.283	0.379
2003	0.390	0.453
2004	0.415	0.375
2005	0.423	0.431
2006	0.407	0.477
2007	0.375	0.400
2008	0.450	0.349
2009	0.406	0.442
2010	0.428	0.425
2011	0.468	0.404

[a] This is the standard deviation of average class grades within courses (of classes that sit the same exam).

Table 12: Heterogeneity by teachers' research skills

	Teaching	Research		
Dep. var	α_{pcy}	n pub	n pub weight	n wp (google)
	[1]	[2]	[3]	[4]
int*post 06* ability q 1	-0.087**	0.146***	0.185**	0.296***
	(0.037)	(0.042)	(0.073)	(0.113)
int*post 06* ability q 2	-0.036	0.200***	0.351***	0.166
	(0.041)	(0.070)	(0.102)	(0.172)
int*post 06* ability q 3	-0.035	0.184	0.197	0.095
	(0.042)	(0.140)	(0.257)	(0.254)
N	3770	6281	6264	6082

Additional controls: age, age squared, all double interactions, teacher fixed effects, year fixed effects. Ability based on tertiles of the teachers fixed effects on research obtained from estimating euqation 8. Robust standard errors clustered by teacher in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 13: Teaching and research skills

	Dep. var= Teaching Fe (θ_p^t)			
	everybody	N>5[a]	everybody	N>5[a]
	[1]	[2]	[3]	[4]
Research FE (θ_p^r)	0.715***	0.795***	0.542***	0.640***
	(0.067)	(0.102)	(0.062)	(0.094)
N	313	109	313	109
Controls	No	No	Yes	Yes

Additional controls: age at entry (linear and squared), gender.
[a] N>5 is referred to the n of observations over which is estimated the teacher fixed effect in the teaching quality regression (for the research quality regression N=10 for every teacher). Robust standard errors in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 14: Sorting research

Dep Variable:	1=exit pre 2006	1=exit post 2006	1=exit pre 2006	1=exit post 2006	1=entry pre 2006	1=entry post 2006
	[1]	[2]	[3]	[4]	[5]	[6]
fe research	-0.062	-0.109***			0.027	-0.039
	(0.087)	(0.038)			(0.049)	(0.053)
fe teachers (year¡2006), research			0.001	-0.071		
			(0.144)	(0.074)		
N	273	260	273	260	289	277

Columns 1, 2, 4 and 5 use fixed effects estimated over the entire time period. Columns 3 and 4 use fixed effects estimated only before the change in incentives. I exclude those exiting because retiring and those entered before 2001. Additional controls (for exit): dummies for year of entry, gender, age at entry, age at entry squared; additional controls (for entry): time trend for year of entry (linear and squared), age at entry, age at entry squared, gender. Bootstrapped standard errors in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 15: Sorting teaching

Dep Variable:	1=exit pre 2006 [1]	1=exit post 2006 [2]	1=exit pre 2006 [3]	1=exit post 2006 [4]	1=entry pre 2006 [5]	1=entry post 2006 [6]
fe teaching	0.123	-0.374***			0.113	-0.309**
	(0.119)	(0.136)			(0.089)	(0.141)
fe teaching (year<2006)			-0.109	-0.086		
			(0.162)	(0.116)		
N	302	283	255	193	302	283

Columns 1, 2, 4 and 5 use fixed effects estimated over the entire time period. Columns 3 and 4 use fixed effects estimated only before the change in incentives. I exclude those exiting because retiring and those entered before 2001. Additional controls (for exit): dummies for year of entry, gender, age at entry, age at entry squared; additional controls (for entry): time trend for year of entry (linear and squared), age at entry, age at entry squared, gender. Bootstrapped standard errors in parenthesis. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table 16: Alternative identification strategies - Research

	Dep. var: Number of publications				
Contr gr	External professors		Bologna professors		
	< m age (43)	> m age (43)	All	Jun	Full
	[1]	[2]	[3]	[4]	[5]
Internal[a]*post 06	0.170*	0.119			
	(0.096)	(0.124)			
Bocconi[a]*post 06			0.162**		
			(0.064)		
Jun bocc[a]*post 06				0.221***	
				(0.080)	
Ord bocc[a]*post 06					0.051
					(0.107)
N	3119	2111	4497	3063	1434

Additional controls: teachers' age and age squared, year fixed effects. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status based on the year 2005.

Table 17: Alternative identification strategy - Teaching

	Dep Var: α_{pcy}			
Contr gr:	External prof		Prof just became tenured	
	< m age (43)	> m age (43)		
	[1]	[2]	[3]	[4]
Internal[a]*post06	-0.061*	-0.034		
	(0.032)	(0.029)		
No full[a]*post06			-0.221***	-0.042*
			(0.052)	(0.025)
N	1958	1931	2068	2068
Teachers fe	Yes	Yes	No	Yes

Additional controls: teachers' age and age squared, dummies for year of arrival at Bocconi and the full interaction on course,degree program, year fixed effects. Columns 3 and 4, only internal teachers (in 2005). Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

[a] Status based on the year 2005.

A Appendix A

Given the exponential utility function and normality of ϵ_i, the agent receives certainty equivalent

$$CE = b_r\alpha_r e_r + b_t\alpha_t e_t + s - \frac{1}{2}(e_r^2 + e_t^2) - \delta e_r e_t - \frac{\eta}{2}(b_t^2\sigma_t + b_r^2\sigma_r) \tag{11}$$

The first order conditions obtained from maximizing the expected utility of the agent with respect to e_r and e_t are:

$$\alpha_r b_r = e_r + \delta e_t; \quad \alpha_t b_t = e_t + \delta e_r \tag{12}$$

and the optimal (internal) solutions are:

$$e_r^* = \frac{b_r\alpha_r - \delta b_t\alpha_t}{1-\delta^2}; \quad e_t^* = \frac{b_t\alpha_t - \delta b_r\alpha_r}{1-\delta^2} \tag{13}$$

Therefore, taking the partial derivatives with respect to b_r, I get:

$$\frac{\partial e_r^*}{\partial b_r} = \frac{\alpha_r}{1-\delta^2} > 0; \qquad \frac{\partial e_t^*}{\partial b_r} = -\frac{\delta\alpha_r}{1-\delta^2} = \begin{cases} > 0 & if \delta < 0 \\ < 0 & if \delta > 0 \end{cases} \tag{14}$$

To show the results stated in Proposition 2, I take the derivatives also with respect to ability:

$$\frac{\partial^2 e_r^*}{\partial b_r \partial\alpha_r} = \frac{1}{1-\delta^2} > 0; \qquad \frac{\partial e_t^*}{\partial b_r \partial\alpha_r} = -\frac{\delta}{1-\delta^2} = \begin{cases} > 0 & if \delta < 0 \\ < 0 & if \delta > 0 \end{cases} \tag{15}$$

B Appendix B

Table B.1: Types of teacher contracts

Description	category
Adjunct Professor	assistant
Researcher Bocconi	assistant
Assistant professor Bocconi	assistant
Assistant Professor (Job Market) Bocconi	assistant
Assistant Professor (Young Foreigners) Bocconi	assistant
1 year scholar Bocconi	assistant
2 year scholar Bocconi	assistant
3 year contract researcher Bocconi	assistant
Assistant professor Bocconi senior	assistant
Researcher Bocconi	assistant
Full contract researcher Bocconi	assistant
Researcher Bocconi on leave	assistant
Associate professor Bocconi	associate
Full Professor Bocconi	full
Extraordinary professor Bocconi	full
Non academics (expert in the subject)	non academics
Associate professor other university	other univ
Associate professor Bocconi on leave	other univ
Temporary contract collaborator SDA[a]	other univ
Collaborator SDA	other univ
Permanent contract collaborator Research centers	other univ
Full contract researcher SDA	other univ
Lecturer SDA	other univ
Full Professor other university	other univ
Full Professor Bocconi on leave	other univ
Associate professor other university	other univ
Full Professor other university	other univ
Researcher other university	other univ
Extraordinary professor other university	other univ
Visiting Professor Long/short Term	other univ

[a] SDA is the Bocconi School of Managers. It offers MBAs and master course only. Faculty is hired and promoted according to different and independent standards.

e professors as external professors.

Table B.2: Random allocation

	Av. final hs grade[a]	Av. female	Av. from Mi	Sd final hs grade
1=int teacher	0.001	0.000	0.001	-0.002
	(0.001)	(0.000)	(0.003)	(0.002)
Teacher's Age	0.000	-0.000	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
1=female teacher	0.001	0.000	0.000	-0.001
	(0.001)	(0.000)	(0.002)	(0.002)
1= course coordin	0.000	-0.001	0.003	-0.000
	(0.001)	(0.001)	(0.003)	(0.002)
N	3889	3889	3889	3889
course*year fe	Yes	Yes	Yes	Yes
F stat joint sign	0.75	0.95	0.39	1.58

The Table plots average classes characteristics (based on students' composition) on teachers' characteristics. Robust standard errors clustered by teacher in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

Table B.3: Step 1: regression on students micro data.

Dependent variable: standardized exam grade	All
	[1]
HS grade	-3.704***
	(0.225)
HS grade^2	4.159***
	(0.131)
1=female	-0.051***
	(0.003)
1=italian	0.142***
	(0.013)
1=from Milan	0.074***
	(0.003)
N	501132

Additional controls: dummies for type of high school, dummies for the full interaction of classes and years (α_{pct}). Robust standard errors clustered by course-year in parentheses. * denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.